Structure and function of endoglucanase V

Gideon J. Davies*, G. Guy Dodson*, Roderick E. Hubbard*, Shirley P. Tolley*, Zbigniew Dauter†, Keith S. Wilson†, Carsten Hjort‡, Jan Møller Mikkelsen‡, Grethe Rasmussen‡ & Martin Schülein‡

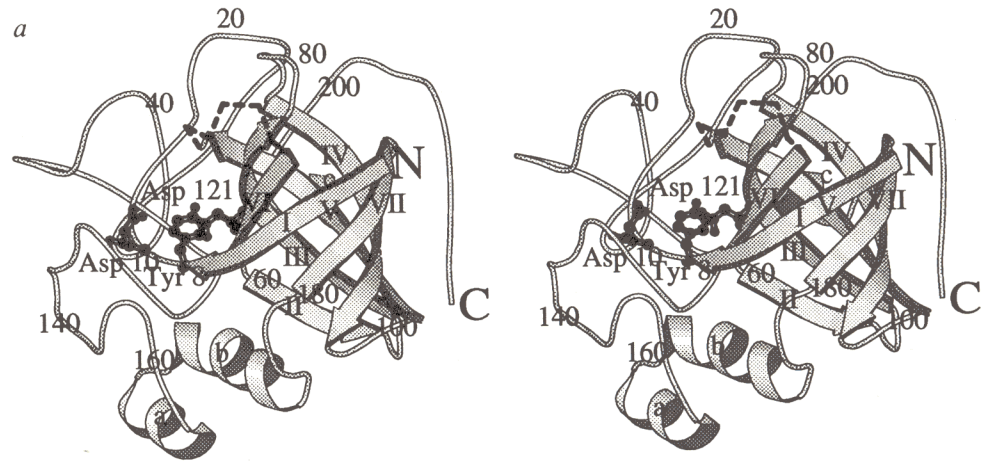
* Department of Chemistry, University of York, Heslington, York YO1 5DD, UK
† EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Germany
‡ Novo Nordisk a/s, Novo Alle, 2880 Bagsvaerd, Denmark

CELLULOSE is the major polysaccharide component of plant cell walls and is the most abundant organic compound on the planet. A number of bacterial[1] and fungal[2] organisms can use cellulose as a food source, possessing cellulases (cellobiohydrolases and endoglucanases) that can catalyse the hydrolysis of the β-(1,4) glycosidic bonds. They can be classified into seven distinct families[3]. The three-dimensional structures of members of two of these families are known[4,5]. Here we report the structure of a third cellulase, endoglucanase V, whose sequence is not represented in any of the above families. The enzyme is structurally distinct from the previously determined cellulases but is similar to a recently characterized plant defence protein[6]. The active site region resembles that of lysozyme, despite the lack of structural similarity between these two enzymes.

Fungal cellulases typically contain a catalytic core linked to a cellulose-binding domain by a flexible peptide segment. The structure of the catalytic core of an endocellulase (EGV) from *Humicola insolens* has been determined by X-ray analysis at 1.6 Å resolution. This enzyme is a new endoglucanase whose sequence has no similarity to any in the original Henrissat classification[3]. The structure has been solved by isomorphous replacement (Table 1). The enzyme has a flattened spheroidal shape with rough dimensions of 42 × 42 × 22 Å. The major structural feature is a six-stranded β-barrel domain. A seventh strand hydrogen bonds to one of the barrel strands but is not part of the barrel itself. The barrel differs from a standard 'jelly-roll' in that it is composed both of anti-parallel and of parallel β-strands. We believe that this represents a new protein topology (Fig. 1) although the barrel core of the structure is similar to that found in 'barwin', a plant defence protein[6] (Fig. 2).

A large, deep groove runs across the surface of the molecule, partitioning the β-barrel from the loop regions. Site-directed mutagenesis has identified two aspartates (10 and 121) critical for activity. They are located on either side of the groove with their Cα atoms some 11.5 Å apart. They are positioned above and either side of a tyrosine residue, Tyr 8, which lies at the bottom of the active site groove. The active site groove contains seven potential subsites for carbohydrate binding. The structure of the catalytic core of EGV is different from both of the two previously reported cellulase structures: the catalytic core of the cellobiohydrolase II (CBH-II) from *Trichoderma reesei*[4] and CelD, a bacterial cellulase from *Clostridium thermocellum*[5]. In contrast to the mixed β-barrel structure of EGV, CBH-II has the α-β-barrel topology first observed in triose phosphate isomerase, the active site being located within an 'enclosed tunnel'. The structure of the bacterial cellulase CelD is again different, having a barrel of 12 α-helices packed alongside an immunoglobulin-type domain.

The similarity between the orientations of the active site residues in CelD and lysozyme has been noted[5]. The active site cleft of EGV also bears a strong similarity to that found both in hen egg white lysozyme (HEWL) and in T4 lysozyme (T4). Catalytic acid groups are found in almost equivalent positions and the shape of the active site clefts are similar although constructed from quite distinct protein folds. One side and the base of the EGV active site groove are primarily made up of four β-strands from the β-barrel. One of these strands contains the catalytic aspartate 121. In HEWL, one side of the active site groove is again provided by three β-strands with the central strand carrying aspartate 52. The similarity in shape between the binding grooves in EGV and HEWL or T4 extends for some 15 Å.



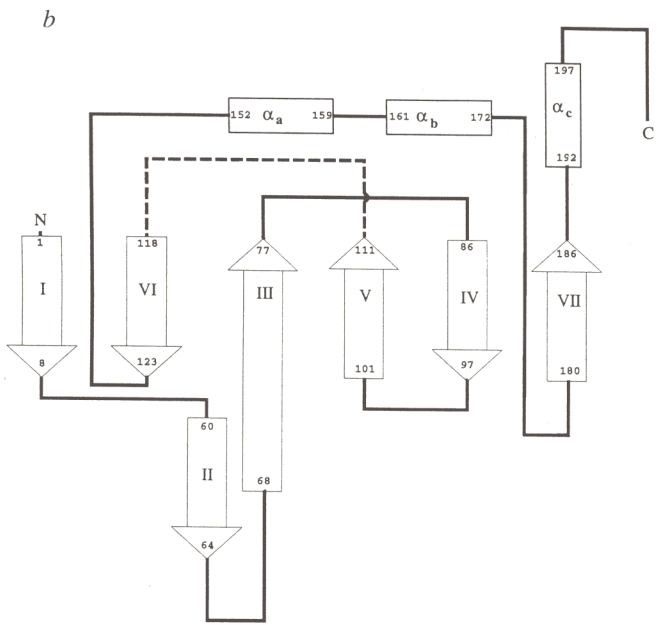


FIG. 1 *a*, Stereo ribbon diagram drawn with the MOLSCRIPT[17] program, showing the overall topology of the EGV molecule. The N- and C-terminal residues are indicated. The flexible loop region (residues 112–117), only seen in substrate/inhibitor complexes, is shown dashed. The catalytic aspartic acid residues and the tyrosine, lying at the floor of the binding-site, are shown in 'ball and stick' representation. Secondary structure was determined using the hydrogen-bonding patterns as defined previously[18], as implemented in the QUANTA program (Molecular Simulation Inc., USA). *b*, Schematic representation illustrating the unusual β-barrel topology of EGV. The flexible loop between barrel strands V and VI is again represented with a dashed line.

TABLE 1 Data collection and phasing statistics

Data collection

Data	Resolution (Å)	R_{merge}*	Number of observations	Number of unique measurements	Completeness (%)
Native	1.6	0.089	102,802	39,421	83
Cellobiose-complex	1.9	0.071	44,074	12,968	97
CH_3HgCl	3.0	0.053	10,271	5.486	75
$LuCl_3$	2.5	0.045	27,987	10,484	84

Phasing statistics

Derivative	R_{deriv}† (%)	R_{cullis}‡			Phasing Power§	Heavy-atom parameters					
		Centric	Acentric	Anomalous		X	Y	Z	B (Å^2)	Occupancy	Anomalous occupancy
						0.025	0.000	0.286	17	0.60	0.54
CH_3HgC1	21.5	0.54	0.66	0.50	1.5	0.040	0.033	0.418	17	0.45	0.45
						0.714	0.239	0.568	28	0.30	0.34
$LuCl_3$	16.5	0.58	0.65	0.60	1.3	0.089	0.094	0.392	29	0.55	0.33
						0.778	0.249	0.979	28	0.70	0.46

Crystals of native EGV were grown by hanging-drop vapour-phase diffusion with 18–22% PEG-8000 as a precipitant. The initial protein concentration was 12 mg ml^{-1} in 10 mM Tris-HCl, pH 8.0. Crystals of EGV are in space group $P2_1$ with cell dimensions $a = 41.8$ Å, $b = 51.3$ Å, $c = 44.4$ Å, and $\beta = 109.6°$. There is one molecule of EGV in the asymmetric unit. Derivative screening using the Hendrix–Lentfer image plate at the EMBL Hamburg outstation, with a Mo sealed tube source, identified two sets of derivatives: methyl Hg chloride (2 mM soak, 5 days) and many salts of the Lanthanide series of metals. Native and derivative data were collected at the Photon factory synchrotron facility using the Weissenberg camera. Data were collected at a wavelength of 1.000 Å to reduce absorption errors and optimize the component of the anomalous scattering from the heavy atoms. $LuCl_3$ (10 mM soak, 10 h) was chosen as, of all the Lanthanide series of metals, it has the greatest values of f'' at low (~1.0 Å) wavelengths. Care was taken with the derivative data to collect Bijvoet pairs on the same image. Screenless Weissenberg data were collected with a camera radius of 286.5 mm, 16° oscillation ranges per image and a coupling constant of 4° mm^{-1}. Data were processed with the WEISS program[10]. All other calculations used the CCP4[11] suite of programs. Phases were calculated using the MLPHARE[12] program with the phases being further improved using SQUASH[13]. The structure was built using the O program[14] and refined using standard techniques with XPLOR[15] and PROLSQ[16]. The crystallographic R-factor is 0.154, at 1.6 Å resolution with r.m.s. deviations from stereochemical target values of 0.010 Å for bonds and 0.025 Å for angles (1–3 bonding distance). Coordinates have been deposited with the Brookhaven Protein Database.

* $R_{merge} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} \langle I \rangle$.

† $R_{deriv} = \sum |F_{deriv} - F_{nat}| / \sum F_{nat}$.

‡ $R_{Cullis} = \sum ||F_{PH_{obs}}| - |F_{PH_{calc}}|| / \sum ||F_{PH_{obs}}| - |F_{P_{obs}}||$ for centric or acentric terms, and
$\sum |\Delta ano_{obs} - \Delta ano_{calc}| / \sum |\Delta ano_{obs}|$ for anomalous differences.

§ The phasing power is the mean value of the heavy-atom structure amplitude divided by the lack of closure error.

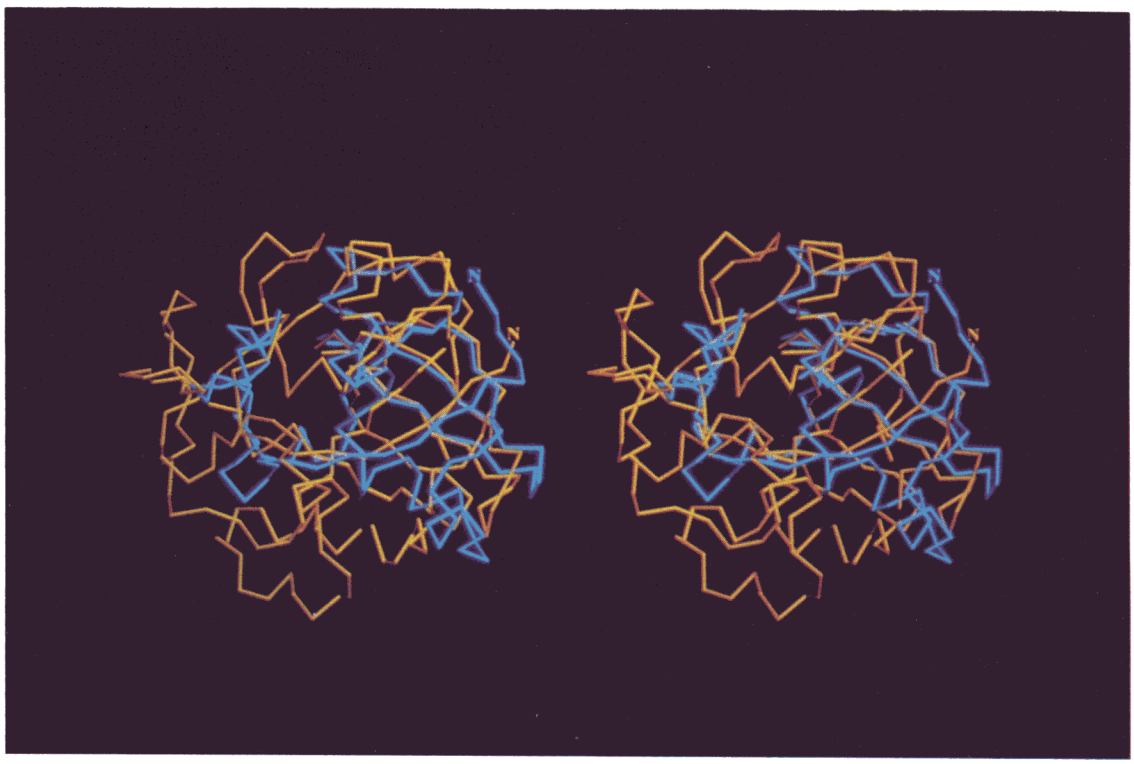


FIG. 2 Stereo figure illustrating the similarity between the β-barrels of 'barwin'[6] (blue) and EGV (yellow). 38 Cα atoms (representing 18% of the EGV Cα atoms) overlap with the barwin barrel with an r.m.s. deviation of 1.56 Å. Coordinates for barwin (and for T4 and hen-egg white lysozyme) were obtained from the Brookhaven Protein Database[19].

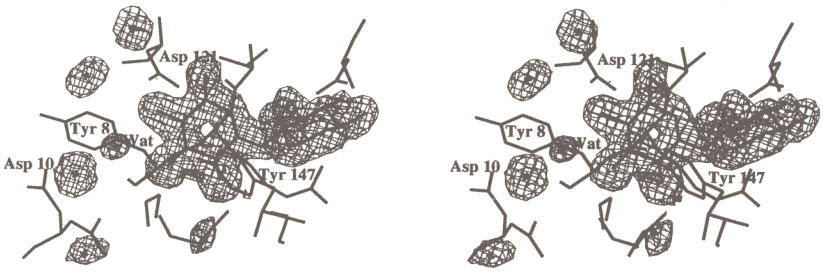


FIG. 3 Stereo figure illustrating the binding of the product cellobiose to EGV. EGV has seven subsites for sugar binding, termed A–G, aiding cleavage of the glycosidic bond between subsites D and E. As with lysozyme[7,8,19] this cleavage is at the reducing end of the substrate. Cellobiose is shown bound to the 'leaving group' sites E and F. A water molecule, labelled 'wat', is bound near Asp 10. Modelling studies using hexasaccharide substrates indicate that, given a small movement (<1 Å), this water molecule would be suitably positioned to make a nucleophilic attack leading to inversion at C1. Crystals of the EGV–cellobiose complex were grown under the same conditions as the native enzyme (Table 1), with the addition of 5 mM cellobiose. The crystals are in space group $P2_1$ with cell dimensions $a=41.75$, $b=56.19$, $c=37.37$ Å and $\beta=93.40°$. The structure was solved by molecular replacement using standard programs[8]. Data were collected to 1.9 Å using an RAXIS-II image-plate and a Cu rotating anode source and processed with the DENZO program (Z. Otwinowski, Yale University). Details of the data quality and completeness are shown in Table 1. The map shown is an F_O-F_c difference map, contoured at a level corresponding to about 0.33e Å^{-3}, from a model consisting of protein atoms only. The current crystallographic *R*-factor is 0.159, with r.m.s. deviations from stereochemical target values of 0.011 and 0.039 Å for bonds and angles, respectively.

This similarity between the lysozyme and EGV active site clefts is not surprising, given that these enzymes bind similar substrates, but it is not seen to this extent in other cellulolytic enzyme structures. Catalysis by lysozyme[7–9] proceeds by an SN_1 mechanism with retention of configuration at C1. The oxycarbonium ion intermediate is stabilized by electrostatic interaction with Asp 52, steric factors dictating that the nucleophilic water molecule is only able to approach the sugar ring once the leaving group has departed. This, together with the participation of Glu 35 in directing the attacking water molecule, results in retention of configuration at C1. In contrast, catalysis by EGV, and by both CBH-II and CelD, proceeds with inversion (C. Schou, B. Henrissat and C. Gey, personal communication). The structure of a cellobiose complex, solved at 1.9 Å, reveals that the two catalytic aspartates are located such that a water molecule bound to Asp 10 could make nucleophilic attack without prior need for product departure (Fig. 3). This stereochemistry dictates that catalysis proceeds with inversion at C1. When cellobiose occupies the 'leaving group' site (subsites E and F), the loop between strands V and VI, disordered in the native structure, interacts with the cellobiose and is well defined.

The completely different folding in the three X-ray structures of cellulases is surprising because we might expect the differences in specificity between these cellulases to be achieved by changes to a basic framework, as is found in other enzyme families such as the ribonucleases, lipases and serine proteases. For those cellulolytic enzymes whose structures are known there is no identifiable relationship in the tertiary folding, but there is a pattern of convergence in their evolution towards similar catalytic sites. The insoluble nature of cellulose as a substrate presents many chemical and metabolic problems for the organism. This may be reflected by the presence of at least eight different families of cellulases and in their multi-domain character. The distinctive structure described here represents the first member of a new family of cellulases. Although the EGV structure is distinct as a cellulase, the presence of the same barrel topology as a polysaccharide-binding (chitin) plant defence protein strongly suggests there is a functional and evolutionary relationship between these two classes of protein.

Note added in proof: The endoglucanase V from *Humicola insolens* has recently been classified into cellulase family 'K' (ref. 20). □

Received 11 January; accepted 21 June 1993.

1. Beguin, P. *et al. Biochem. Soc. Trans.* **20,** 42–46 (1992).
2. Wood, T. M. *Biochem. Soc. Trans.* **20,** 46–53 (1992).
3. Henrissat, B. *Biochem. J.* **280,** 309–316 (1991).
4. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. & Jones, T. A. *Science* **249,** 380–386 (1990).
5. Juy, M. *et al. Nature* **357,** 89–91 (1992).
6. Ludvigsen, S. & Poulson, F. M. *Biochemistry* **31,** 8783–8789 (1992).
7. Blake, C. C. F. *et al. Proc. R. Soc.* B**167,** 365–377 (1967).
8. Phillips, D. C. *Proc. R. Acad. Sci.* **57,** 484–495 (1967).
9. Strynadka, N. C. J. & James, M. N. G. *J. molec. Biol.* **220,** 401–424 (1991).
10. Higashi, T. *J. appl. Crystallogr.* **22,** 9–18 (1989).
11. CCP4, The SERC (UK) Collaborative Computing Project No. 4: *A Suite of Programs for Protein Crystallography* (Daresbury, UK, 1979).
12. Otwinowski, Z. in *Proc. CCP4 Study Weekend* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–85 (CCP4, Daresbury, UK, 1991).
13. Zhang, K. Y. J. & Main, P. *Acta Crystallogr.* A**46,** 377–381 (1990).
14. Jones, T. A., Zou, J-Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* A**47,** 110–119 (1991).
15. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
16. Hendrickson, W. A. & Konnert, J. H. in *Structure, Conformation and Evolution* Vol. 1 (ed Srinivasan, R.) 43–57 (Pergamon, Oxford, 1981).
17. Kraulis, P. J. *J. appl. Crystallogr.* **24,** 946–950 (1991).
18. Richardson, J. *Adv. Prot. Chem.* **34,** 167–339 (1981).
19. Bernstein, F. C. *et al. J. molec. Biol.* **112,** 535–542 (1977).
20. Henrissat, B. & Bairoch, A. *Biochem. J.* **293,** 781–788 (1993).

ACKNOWLEDGEMENTS. We thank N. Sakabe for provision of data collection facilities at the Photon Factory synchrotron facility, H. Savage and T. Wilkinson for discussions, M. Hartshorn for assistance with the figures and A. Murzin (LMB, Cambridge) for pointing out the similarity between EGV and barwin. We thank the SERC and Novo-Nordisk a/s for financial support.